Where Can I Help? Human-Aware Placement of Service Robots

Lilli Bruckschen   Kira Bungert   Moritz Wolter   Stefan Krumpen
Michael Weinmann   Reinhard Klein   Maren Bennewitz

Abstract— As service robots are entering more and more homes it gets evermore important to find behavior strategies that ensure a harmonic coexistence between those systems and their users. In this paper, we present a novel approach to enable a mobile robot to provide timely assistance to a user moving in its environment, while simultaneously avoiding unnecessary movements as well as interferences with the user. We developed a framework that uses information about the last object interaction to predict possible future movement destinations of the user and infer where they might need assistance based on prior knowledge. Given this prediction, the robot chooses the best position for itself that minimizes the time until assistance can be provided as well as avoids interferences with other activities of the user. We evaluated our approach in comparison to state-of-the-art methods in simulated environments and performed a user study in a virtual reality environment. Our evaluation demonstrates that our approach is able to decrease both the time until assistance is provided and the travel distance of the robot as well as increases the average distance between the user and the robot in comparison to state-of-the-art systems. Additionally, the robot behavior generated by our method is rated as more pleasant by our study participants than comparable literature approaches.

Fig. 1: Motivation of our approach. While the user moves to a microwave, the robot predicts it can offer service at the table by providing a cup.

I. INTRODUCTION

Assisting humans in everyday tasks is an important field of robotics. The influence of this field will likely grow during the next years as the demographic makeup of almost all western societies is changing [1]. As the baby boomer generation grows old and retires, it seems unlikely that demand for private and retirement home assistance can be met by an increase of the human workforce alone. Possible tasks include getting in and out of bed, fetch the walking cane, getting in and out of a wheelchair, or fetching an object from an elevated board.

To efficiently assist people in such scenarios, robots must be able to anticipate where a person will need help in the future and where not. Both aspects are equally important, as they allow the robot to timely provide help if needed and avoid scenarios where it would disturb the users, e.g., by unnecessary movements into areas where no assistance is required. Fig. 1 shows an example scenario for such a human-robot cooperation.

In this paper, we address these challenges by presenting a novel anticipation approach that uses a prediction system to infer likely future movement goals of the human, estimates the likelihood that the user will need assistance at those

locations, and computes a non-disturbing robot position close to the most likely next locations where the robot can help the user. To accomplish this, we developed a probabilistic framework to predict a sequence of future navigation goals of the human based on our previously presented object-interaction prediction system [2] and taking into account the user's observed pose and velocity. We further propose to learn a model to infer at which of the possible next navigation goals the user will need the robot's help. To realize this, we associate each object with a navigation goal and categorize it based on the capabilities of the deployed robot and the respective object-specific assistance. Given these information, we calculate a robot position between the predicted likely next destinations at which the user will need help, considering their probabilities. We evaluated the robot behavior generated by our approach in comparison to state-of-the-art techniques [3], [2], [4] in simulated environments with respect to the time until assistance can be provided, traveled distance, and average distance to the user. In addition, we conducted a virtual reality user study with 20 participants to infer how humans perceive the robot behavior resulting from our approach compared to other systems.

II. RELATED WORK

As service robots become more common research in their regards increases likewise. In a recent study, Tussyadiah and Park found that, besides familiarity with humans in form of anthropomorphic design and adherence to social norms, the perceived intelligence of the robot is of great importance for the user [5]. It can be increased by correctly interpreting the intentions of a user and acting based on this

information. This often corresponds to predicting either the user's next navigation goal or activity. For example, Ye *et al.* used a hidden Markov model to predict the category of the user's next activity (e.g., food, shopping, entertainment) and location where this activity is likely to occur [6]. Best and Fitch applied a Bayesian framework to estimate the navigation goal and future trajectory of an agent in a static environment [7]. Bayoumi and Bennewitz tackled the same problem but developed a Q-learning method that utilizes typical human trajectories on a known map to predict the user's navigation goal as well as their trajectory [4]. The authors focused on minimizing unnecessary movements of the robot, e.g., the robot should not follow the human if they take a detour to their navigation goal. In contrast, we focus on object interactions to predict future navigation goals, i.e., our approach is, in general, independent of a specific environment or map and can infer where the user might need the help of the robot. Carlson *et al.* demonstrated how intention prediction can directly be used by robotic systems to help humans [8]. The authors designed an intelligent wheelchair that is able to predict the user's movement intention and helps to reach their navigation goal.

A promising development in this field is the utilization of smart home systems to monitor human-object interactions even if they are not directly observable through the robot. For example, Amri *et al.* used a multi-modal system to monitor activities of elderly people [9]. Muztoba *et al.* proposed special input devices with which a user can inform the robot about new commands at any location inside a smart home [10]. The authors experimented with gesture and speech detection as well as brain-machine interfaces. Furthermore, Alam *et al.* demonstrated how a smart home system could be used in combination with a Markov model to accurately predict future activities of humans based on the observation history [11]. Our framework would also benefit from the data collected in a smart home and could be easily deployed in such environments.

III. HUMAN-AWARE SERVICE ROBOT PLACEMENT

The goal of our work is to enable a mobile robot to predict where a moving user will next need assistance and place itself at an appropriate location without causing disturbances to the user. Our framework relies on the observation that knowledge about human-object interactions is essential for the prediction of both, the movement of humans, as well as possible objects where assistance may be needed. In our previous work [2], we showed that knowledge about the last object a human has interacted with enables a robot to infer with which object they will interact next and predict the navigation goal of a moving human. In this work, we build upon this framework and extend our system to predict, first, how humans move over a longer time horizon and, second, where they will likely need help. To do so, our approach learns for the objects whether the user may need the robot's help while interacting with them. Using the prediction about the user's movements and activities as well as the information whether the robot's assistance will be needed, we compute a

utility map to find the best position where the robot should place itself. In summary, our framework consists of four components:

A. Gathering of prior knowledge about subsequent object interactions and at which objects the robot's assistance is desired.

B. Estimation of the user's state, i.e., orientation, location, and velocity, as well as human-object interactions detected from the robot's observations.

C. Prediction of future object interactions and corresponding navigation goals based on the prior knowledge and the user's current state using Bayesian inference.

D. Calculation of a utility map using likely future positions at which the user needs the robot's help, according to the prior knowledge, and determining the position with the highest utility.

A. Prior Object Knowledge

To model the user's transition probabilities between interacting with different objects, we use a distribution called interaction model I [2]. This model encodes the probability $I(\tau_B|\tau_A)$ that a user who has last interacted with an object of class τ_A will next interact with an object of class τ_B. The interaction model we apply in this work was prelearned based on 195 human-object interactions, a part of our dataset is published as the Bonn activity maps dataset [12]. We used 17 different object classes which we assigned to 4 activity classes to reduce complexity and allow a general classification of whether help can be provided at an object: resting *(bed, wardrobe, sofa)*, food processing *(bottles, cups, microwaves, workbenches, refrigerators, coffee machines)*, office work *(chairs, tables, laptops, whiteboards, cupboards)*, and hygiene *(toilets, washbasins, bathtubs)*. We performed eight interviews with students from the University of Bonn about preferable service robot behavior. In this context, we asked participants at which activities they would like the robot to provide assistance. Based on the results, we concluded that our robot should provide assistance for the user during resting, e.g., by providing fetch tasks for the user, and food processing activities. Note that in this work, we do not assume that the robot actually carries out the assistance actions but focus on human activity prediction and human-aware robot placement.

B. Human State and Robot Observations

We model the state of the human as $S := (\mathcal{X}_h, \theta, \nu)$, with $\mathcal{X}_h$ as their position in a grid map of the environment, θ as their orientation, and ν as their velocity at observation time. $\mathcal{X}_h$ and ν can be estimated using the robot's onboard sensors in combination with the known position of it on the map. Additionally, we analyze RGB-D data and infer the torso orientation θ by combining estimations about the joint positions from the 2D image and depth data. Similar to Biswas *et al.* [13], we obtain a pixelwise joint position probability map by applying an implementation of OpenPose [14]. We then estimate the torso normal by analyzing the shoulder and hip key points. For this purpose,

we compute the cross product for each edge of the rectangle formed by the shoulder and hip joints. By computing the orientation in this way, we achieve a more precise orientation estimation compared to our previous system [2], which relied on a rough segmentation into four regions, each with a size of $90°$. To detect human-object interactions, we used a detection system based on RGB-D and body pose data [15]. An object interaction is registered when the human faces an object and places at least one hand in close proximity to it.

C. Prediction of Future Object Interactions

Let M be the grid map of the environment. We use Bayesian inference to infer future human-object interactions, with the previously learned interaction model I as prior knowledge and the current human state $S := (\mathcal{X}_h, \theta, \nu)$ as observation. Let $bel(o_i^n) = P(o_i^n|S)$ be the belief that the user's n-th future interaction will be with object $o_i = (\mathcal{X}_{o_i}, \tau_{o_i})$ at position $\mathcal{X}_{o_i}$ on M and object class τ_{o_i} given the observed user state S. The probability of object interactions with $n > 1$ can than be recursively inferred using I as

$$bel(o_i^n) = \sum_{o_j^{n-1}} P(o_i^n|o_j^{n-1}, S) \cdot P(o_j^{n-1}|S) \qquad (1)$$

$$= \sum_{o_j^{n-1}} P(o_i^n|o_j^{n-1}, S) \cdot bel(o_j^{n-1}) \qquad (2)$$

$$= \sum_{o_j^{n-1}} P(o_i^n|o_j^{n-1}) \cdot bel(o_j^{n-1}) \qquad (3)$$

$$= \sum_{o_j^{n-1}} I(\tau_{o_i^n}|\tau_{o_j^{n-1}}) \cdot bel(o_j^{n-1}) \qquad (4)$$

using the law of total probability (Eq. (1)), recursion (Eq. (2)), the fact that o_j^n is independent of the current human state S given o_j^{n-1}, and the definition of the interaction model (Eq. (4)).

The interaction at $n = 1$ can be inferred with the current observation about the human state and the prior knowledge:

$$bel(o_i^1) = \frac{P(S|o_i^1)P(o_i^1)}{\sum_{o_j^1 \in O} P(S|o_j^1)P(o_j^1)} \qquad (5)$$

$$= \eta \cdot P(S|o_i^1)P(o_i^1) \qquad (6)$$

$$= \eta \cdot P(S|o_i^1)I(\tau_{o_i^1}|\tau_{o^0}) \qquad (7)$$

using Bayes's rule (Eq. (5)), a normalization constant (Eq. (6)), and the definition of the interaction model (Eq. (7)), with τ_{o^0} as the class of the last observed object the human interacted with. If o^0 is unknown we obtain $I(o_i^1|o^0)$ using marginalization over all possible previous objects.

The likelihood $P(S|o_i^1)$ considers the user's orientation and distance to the object. Let $\mathcal{P}_{\mathcal{X}_h \to \mathcal{X}_{o_i^1}}$ be the A* path from the position of the human $\mathcal{X}_h$ to the position of the object o_i^1. Let further $\Delta a(\theta, \theta_{opt})$ be the difference beween the human's

orientation θ and the orientation θ_{opt} they would have if they moved to the next position on the A* path $\mathcal{P}_{\mathcal{X}_h \to \mathcal{X}_{o_i^1}}$.

Let $\Delta t(\mathcal{X}_h, \mathcal{X}_{o_i}, \nu)$ be the time the human would take from their current position to $\mathcal{X}_{o_i}$ on $\mathcal{P}_{\mathcal{X}_h \to \mathcal{X}_{o_i}}$ with respect to their observed velocity ν. To decrease the likelihood of objects from which the user moves away and to model the fact that they cannot turn around spontaneously, we consider the distance the user would travel until the next observation update if $\Delta a(\theta, \theta_{opt}) > 180°$. In other words, if the user moves away from an object, we assume that they cannot turn around before the next observation update is scheduled and take this into account when computing Δt, which is defined as follows:

$$\Delta t(\mathcal{X}_h, \mathcal{X}_{o_i}, \nu) = \begin{cases} \frac{dist(\mathcal{X}_h, \mathcal{X}_{o_i})}{\nu}, \text{ if } \Delta a(\theta, \theta_{opt}) < 180° \\ \frac{dist(\mathcal{X}_h, \mathcal{X}_{o_i}) + (\nu \cdot f_{update}^{-1})}{\nu}, \text{ else} \end{cases}$$
$$(8)$$

with $dist(\mathcal{X}_h, \mathcal{X}_{o_i})$ as the A* distance between the position of the human $\mathcal{X}_h$ and the position of the possible goal object $\mathcal{X}_{o_i}$ and $f_{update} \left[\frac{1}{s}\right]$ as the update frequency of the prediction.

The smaller the Δa and Δt, the higher the likelihood and we therefore define the observation likelihood as

$$P(S|o_i^1) = \Delta a(\theta, \theta_{opt})^{-1} \cdot \Delta t(\mathcal{X}_h, \mathcal{X}_{o_i}, \nu)^{-1}. \qquad (9)$$

Thus, we have defined all components to compute the belief about the user's n-th future object interaction and, thus, navigation goal.

D. Calculating the Maximum Utility Position for the Robot

By combining the prior knowledge and the prediction, the robot is able to estimate at which locations the user will likely need its help in the future. However, the robot must still decide where it should place itself between these positions. This is a non-trivial task as the prediction is probabilistic and equally likely goal objects may be far apart. It is also not a priori clear how far the robot should look into the future to make effective predictions. To estimate this, we evaluated the belief over different values of n, which denotes the n-th future object interaction or navigation goal, in regards to the standard deviation of the probability of the different goals. A low standard deviation corresponds to a scenario in which no strong prediction can be made as the goals are more or less equally likely. Therefore, we are interested in values of n with a relatively high average standard deviation. Thus, we evaluated the standard deviation for different n in our simulated environments (which are described in Sec. IV-A). The results are shown in Fig. 2. As can be seen, the average standard deviation stagnates at a low level after $n = 4$. Therefore, we consider only the next four goal objects of the human to decide on the best robot position.

When required, the robot should provide assistance to the human as soon as possible. Therefore, the robot should prioritize promising goal positions at a low n value over such with a higher n value. To ensure this we use a probabilistic
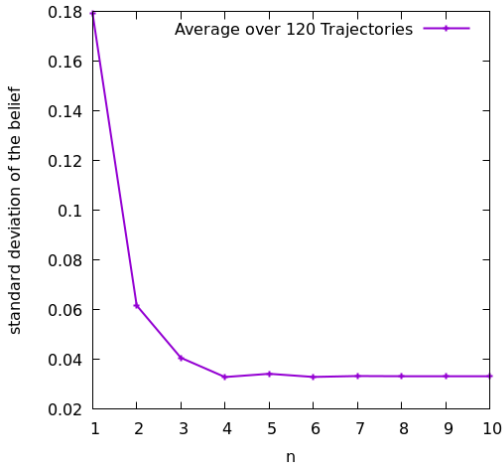
Fig. 2: Standard deviation of the belief of possible goal locations (corresponding to the goal objects) for different values of n. As can be seen, there are no major differences for prediction results with n values greater than 4.

weight function $w(n)$ for the probability that the human will need the robot's help at the given value of n. This function is defined as the number of significantly likely goal objects at which the human would need assistance divided by the number of all significantly likely goal objects, both for the given value of n. Here, we define a goal object as significantly likely if its goal probability is above the sum of the average goal probability and their standard deviation for the given value of n. In other words, a high value of $w(n)$ corresponds to a high probability that the human will need assistance at the given value of n, whereas a low value corresponds to a low likelihood of this event. Using this function as well as the knowledge at which objects the human needs the robot's help, we can compute the best robot position between the possible goal locations weighted by the probability that the human will arrive at them at time step n, using $w(n)$ and the corresponding belief distribution. Let max be the maximum value for n (in our case 4), $dist(\mathcal{X}, \mathcal{X}_{o_j})$ be the A* distance between $\mathcal{X}$ and the position $\mathcal{X}_{o_j}$ of the object o_j and h be a function that returns 1 for objects where the robot should provide assistance and 0 otherwise, based on the prior knowledge (in our case h returns 1 for objects of the resting and food processing class, as defined in Sec. III-A). The utility $U^n(\mathcal{X})$ of position $\mathcal{X}$ considering the predicted n next object interactions is therefore computed as follows:

$$U^n(\mathcal{X}) = w(n) \cdot \sum_{o_j} \frac{bel(o_j^n) \cdot h(o_j)}{dist(\mathcal{X}, \mathcal{X}_{o_j})} + (1\text{-}w(n)) \cdot U^{n+1}(\mathcal{X}) \tag{10}$$

for $0 < n < max$ and else

$$U^{max}(\mathcal{X}) = w(n) \cdot \sum_{o_j} \frac{bel(o_j^{max}) \cdot h(o_j)}{dist(\mathcal{X}, \mathcal{X}_{o_j})} \tag{11}$$

Fig. 3 shows an example of a computed utility map in a simulated environment.
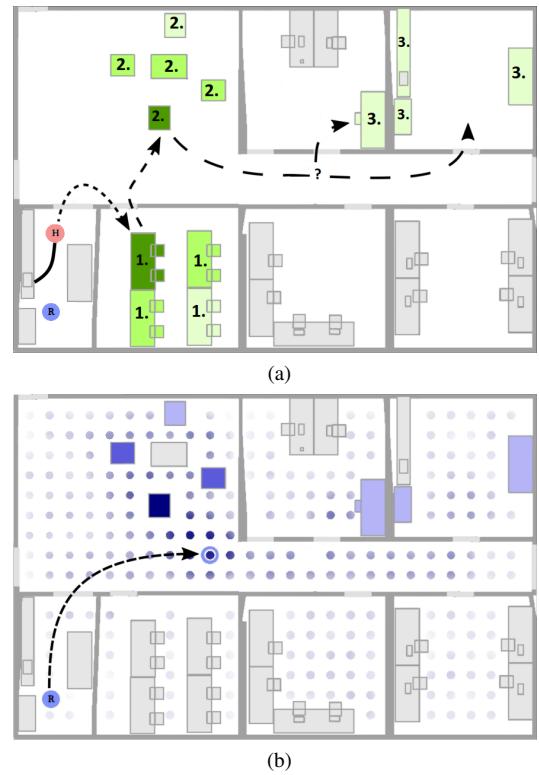

Fig. 3: (a) Predicted next object interactions of the user. The robot (blue) observes the human (pink) interacting with a coffee machine and moving away from it (black line). Using our prediction framework, the robot computes the likely next objects with which the human will interact. Likely goals are depicted in green with their n value (corresponding to the n-th object interaction), the darker the color the higher the likelihood. The next two objects can be predicted with a certain accuracy, the third, however, is highly unsure as 4 different objects are likely. (b) Utility map based on the predictions computed in (a). Objects at which the robot can assist the human are colored blue, as are possible robot positions. The darker the color the higher the goal belief and utility, respectively. Once the utility map is calculated, the robot computes a path to the position with the highest utility (blue circle).

Once the position $\mathcal{X}_{max}^n$ with the highest utility for the n-th next object interactions is determined, the robot moves towards it, following the path computed by A* on the grid map, until the user's goal position is updated or the human calls for assistance.

IV. EXPERIMENTAL EVALUATION

We evaluated our approach in a simulated environment with sampled human trajectories and compared the results to our previous work as well as to two other state-of-the-art approaches with respect to the time the robot would take to reach the human at objects at which they need the robot's assistance, the total distance the robot traveled, and the distance to the user. To measure how users rate the robot behavior generated by our approach in a controlled environment, we performed a virtual reality experiment.

A. Quantitative Evaluation

For the quantitative evaluation in simulation, we used 8 environments with sizes between $100\,m^2$ and $150\,m^2$,

	Avg. time until assistance [s]	Avg. distance to human [m]	Avg. distance traveled [m]
Our approach	**0.63**	**19.83**	**83.90**
Predictive approach [2]	3.69	8.54	172.81
Q-learning approach [4]	5.94	6.90	151.98
Follower approach [3]	2.61*	3.23	144.16

TABLE I: Evaluation results of our approach compared to state-of-the-art methods. As can be seen, our approach outperforms all other methods. With the exception of the avg. time until assistance between the follower approach and our approach (marked with *), all results are statistically significant at $p < 0.05$. Thereby, we achieve in comparison a more efficient and less disturbing navigation strategy than other state-of-the-art methods with a comparable arrival time to a follower approach.

a grid resolution of 0.25 meter and 110 different objects from 17 different classes using the V-REP editor [16]. As prior knowledge we used the objects and interaction model discussed in Sec. III-A.
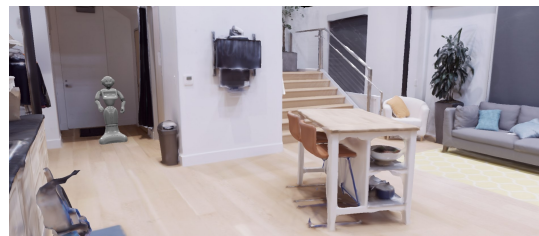
We evaluated the approaches using 540 test trajectories distributed over all environments. These were randomly generated, based on a training set of 128 previously recorded object-interaction sequences. The same set of test trajectories was used for all evaluations and provides the ground truth information about the user's navigation goal.

We compared our approach with state-of-the art navigation approaches that try to minimize the time until a robot arrives at a destination at which help is needed, once the user required help. These are a Q-learning based approach by Bayoumi *et al.* [4], our previous predictive approach [2], which does neither consider interactions further in the future nor differentiates between objects where the human may or may not need help, and a follower approach which does not use movement predictions by Tee *et al.* [3]. As metrics we used the time the robot would take to reach the user at objects at which they need the robot's assistance, the average distance between the robot and the user, and the total distance the robot traveled.
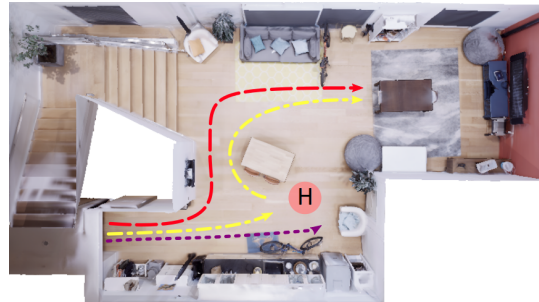
As it is our goal to provide timely assistance to the user while minimizing interference with other activities, an optimal result would be a low value for the time until assistance is provided, a high distance to the human, and a low total distance traveled. The results of the evaluation are depicted in Tab. I. As can be seen, our approach outperforms all other approaches on the dataset. We performed paired t-tests with $\alpha = 0.05$ on the results and found that with the exception of avg. time until assistance of the follower approach our approach was significantly better than all other approaches in all metrics. This was expected as the follower approach stays right behind the human, while all other approaches try to move to the goal of the user, thereby possibly computing a false positive location. However our approach still performs statistically similar than the follower approach on this metric.

B. Evaluation in a Virtual Reality Setting

To evaluate how humans rate the robot behavior generated by our approach, we conducted a virtual reality experiment

(a)

(b)

Fig. 4: (a) View of the user in the virtual reality environment to evaluate different robot navigation behaviors. (b) Bird's eye view of the complete environment. The non-moving participant (pink circle) observed three different trajectories, corresponding to the robot behavior resulting from our approach (red trajectory), the prediction approach presented in [2] (yellow trajectory), and the follower approach presented in [3] (violet trajectory).

with 20 student participants from the University of Bonn, using a HTC Vive. We used an environment from the Facebook replica dataset [17] and a model of the Pepper robot. The position of the participants was in front of a table and we instructed them to observe the behavior of the robot while staying at their position. We showed each participant three different robot trajectories in the VR environment, one after the other, as shown in Fig. 4. Participants were encouraged to mention any robot behavior that they did not like during the experiment. First, we applied our presented approach (red line), here the robot is aware of the user's intentions and moves foresightedly. Accordingly, the robot predicts that the user does not need help at their current position and moves close to the likely next point where the human is expected to need help. With the approach presented in [2], the robot does not know where the user might need its help and checks if they require assistance at the predicted immediately next navigation goal. As the robot is not called to help the user, it realizes that the human does not need assistance at this location and moves to the predicted next navigation goal based on the observed object interaction (yellow line). The last trajectory (purple line) results from the follower approach [3]. Here, the robot moves to the position of the user. However, in contrast to the previous approaches, the robot does not move in advance to the user's predicted next navigation goal but stays close to the human. We didn't evaluate the Q-learning approach in the VR experiment as no training trajectories were available for this map [4].

After all scenarios, the participants were asked to rate

	pleasant	neutral	unpleasant
Our approach	**0.85**	0.15	0.00
Predictive approach [2]	0.05	0.25	**0.70**
Follower approach [3]	0.10	**0.50**	0.40

TABLE II: Results of the virtual reality experiment evaluating the robotic navigation behavior of three different approaches. Each navigation behavior could be rated as *pleasant*, *neutral*, or *unpleasant*. In general, participants rated our approach as pleasant, while the approach [2], which only predicts the next navigation goal, was mostly rated as unpleasant for situations in which the robot's help is not needed. The follower approach [3] was rated as neutral but too close to the user and, therefore, unpleasant.

the different robot navigation behaviors based on their satisfaction with them. The results are depicted in Tab. II. As can be seen, 85% of the participants rated our approach as pleasant, while 15% rated it as neutral. The behavior of the robot was described as clear and predictable, participants were also pleased with the robot's distance to them. The second approach [2], was rated as unpleasant by 70%, while 25% rated it as neutral, and only 5% as pleasant. Participants were unsure about the robot's intentions, as the robot first moved towards them and then moved away from them. Some were even annoyed by this behavior. The follower approach was rated as unpleasant by 40%, as neutral by 50%, and as pleasant by 10%. The biggest critique of the users was that the robot came too close to them. However, most of the participants also noted that the robot acted in a way they expected it to. These results support our assumption that humans feel more comfortable if the robot avoids unnecessary interference or coming too close to the user. We also observed that humans highly appreciate predictability in the robot's behavior, as the most unpleasant approach was mainly criticized for its unpredictable navigation behavior and the third, while unpleasantly close to the user, praised for its simple navigation strategy. This may present a limitation for our approach as false positive movement predictions would most likely result in unpredictable robot movements for the human. We will take this into account for our future work.

V. Conclusions

In this paper, we presented a novel approach to compute an optimal position for a robotic assistant based on the user's likely future object interactions for which they might need the robot's help. We accomplished this by combining prior knowledge about typical human-object interactions, observations about the human, and Bayes' inference to predict possible future navigation goals. Using these predictions, we compute a utility map by weighting likely future goal positions at which the human may need help. Using this map, we find the robot position which minimizes the time the robot needs to provide assistance if needed, as well as avoids unnecessary movements of the robot and disturbances of other activities of the user. As demonstrated by our experimental evaluation, our approach significantly outperforms state-of-the-art methods in terms of distance to the human during travel and average distance traveled, thereby also decreasing the risk of potential disturbances. Simultaneously it performs statistically comparable to a follower approach in regards to time until help can be provided by the robot. Our approach was further rated as more pleasant than both, a state-of-the-art prediction method as well as a state-of-the-art follower system, by increasing the distance to the human and being considered as predictable. We thereby achieved our goal to create a human-aware placement approach for service robots, which achieves a low time until assistance can be provided and is rated favorably by human users.

References

[1] W. Lutz, G. Amran, A. Blanger, A. Conte, N. Gailey, D. Ghio, E. Grapsa, K. Jensen, E. Loichinger, G. Marois, R. Muttarak, M. Potanokov, P. Sabourin, and M. Stonawski, "Demographic scenarios for the EU – Migration, population and education," 2019.

[2] L. Bruckschen, N. Dengler, and M. Bennewitz, "Human Motion Prediction Based on Object Interactions," in *Proc. of the European Conference on Mobile Robots (ECMR)*, 2019.

[3] M. Tee Kit Tsun, B. Lau, and H. Siswoyo Jo, "An improved indoor robot human-following navigation model using depth camera, active ir marker and proximity sensors fusion," *Robotics*, vol. 7, no. 1, 2018.

[4] A. Bayoumi and M. Bennewitz, "Learning optimal navigation actions for foresighted robot behavior during assistance tasks," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2016.

[5] I. P. Tussyadiah and S. Park, "Consumer evaluation of hotel service robots," in *Information and communication technologies in tourism.* Springer, 2018.

[6] J. Ye, Z. Zhu, and H. Cheng, "What's your next move: User activity prediction in location-based social networks," in *Proceedings of the SIAM International Conference on Data Mining.* SIAM, 2013.

[7] G. Best and R. Fitch, "Bayesian intention inference for trajectory prediction with an unknown goal destination," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[8] T. Carlson and Y. Demiris, "Human-wheelchair collaboration through prediction of intention and adaptive assistance," in *IEEE International Conference on Robotics and Automation*, 2008.

[9] M.-H. Amri, Y. Becis, D. Aubry, and N. Ramdani, "Indoor human/robot localization using robust multi-modal data fusion," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2015.

[10] M. Muztoba, R. Voleti, F. Karabacak, J. Park, and U. Y. Ogras, "Instinctive assistive indoor navigation using distributed intelligence," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 23, no. 6, 2018.

[11] M. R. Alam, M. B. I. Reaz, and M. M. Ali, "Speed: An inhabitant activity prediction algorithm for smart homes," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 4, 2011.

[12] J. Tanke, O.-H. Kwon, P. Stotko, R. A. Rosu, M. Weinmann, H. Errami, S. Behnke, M. Bennewitz, R. Klein, A. Weber, *et al.*, "Bonn activity maps: Dataset description," *arXiv preprint arXiv:1912.06354*, 2019.

[13] A. Biswas, H. Admoni, and A. Steinfeld, "Fast on-board 3d torso pose recovery and forecasting," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2019.

[14] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," *arXiv preprint arXiv:1812.08008*, 2018.

[15] L. Bruckschen, S. Amft, J. Tanke, J. Gall, and M. Bennewitz, "Detection of Generic Human-Object Interactions in Video Streams," in *Proc. of the International Conference on Social Robotics (ICSR)*, 2019.

[16] M. F. E. Rohmer, S. P. N. Singh, "V-REP: A versatile and scalable robot simulation framework," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[17] J. Straub, T. Whelan, L. Ma, Y. Chen, *et al.*, "The Replica dataset: A digital replica of indoor spaces," *arXiv preprint arXiv:1906.05797*, 2019.